

# DSP Based PMD Emulators for Built-in Testing of Coherent Optical Receivers

A. Gokcen Mahmutoglu, Alper T. Erdogan and Alper Demir

**Abstract**—We propose discrete-time polarization mode dispersion (PMD) models that are compatible with the emerging coherent receiver techniques, and statistical sampling schemes for the model parameters. These models use multiple-input multiple-output (MIMO) finite impulse response (FIR) filters that are lossless and therefore lend themselves as perfect candidates for emulation of fiber channels suffering from PMD without polarization dependent loss (PDL). The concatenated composition of these filters resembles the continuous time lumped model of PMD channels and offers a flexible emulator and compensator structure in terms of computational complexity which constitutes the main bottleneck for real-time DSP applications.

The parameter sampling problem for accurate approximation of PMD channels considering their statistical behavior is tackled using three different approaches, which we introduce in order of decreasing deviation from the desired statistics and increasing computational complexity. We present simulation results for each sampling method and compare them with the desired statistics.

## I. INTRODUCTION

Polarization mode dispersion (PMD) is an important concern in the design of high-speed optical fiber communication systems. Although recent advances in coherent receiver technologies [1], [2], [3] brought some relief from optical distortions in optical fibers with the use of compensation algorithms, PMD continues to be a major impediment at bit/symbol rates of 100 Gb/s and more [4], [5]. PMD results in a random coupling and a speed difference between the two, normally degenerate, orthogonal polarization modes of propagation in single-mode optical fibers due to random imperfections breaking their circular symmetry, and therefore has a statistical nature [6]. These effects result in pulse broadening and intersymbol interference (ISI) as well as the mixing of two orthogonal polarization channels which limit the transmission speed while increasing the system outage probability.

Accurate emulation of PMD is required for the development and testing of optical fiber communication systems. Currently, development of such systems is based on testing procedures employing conventional PMD emulators built with optical and mechanical components. These emulators can be constructed with different techniques such as simple polarization beam splitter (PBS) and phase retardation elements [7] which only emulate first-order PMD, transmission loops with computer operated polarization controllers and polarization maintaining fibers (PMF) [8] and a concatenation of a high number of birefringent wave plates [9]. The common property of these methods is that they either have limited emulation

capability and programmability or suffer from high structural and computational complexity. Moreover, the incorporation of optical elements such as polarization scramblers and rotational mechanical systems results in expensive devices that are also bulky in size.

State of the art transceiver technologies make use of coherent receiver techniques and digital signal processing (DSP) algorithms in order to increase spectral efficiency and system capacity and compensate for the shortcomings of the legacy technology such as low tolerance for PMD at data rates higher than 40 Gb/s [10]. Furthermore, as counterparts of PMD emulators with optical components, PMD compensators used in direct detection receiver schemes suffer from the same inconveniences of complexity and bulkiness. In this paper, we propose PMD emulation methods based on a discrete-time scheme that is compatible with the emerging coherent receiver techniques. We develop algorithms that can be used on custom chips already present in high speed coherent receivers with minimal alterations of these architectures or on ordinary digital processors for impairment assessment methods relying on off-line data processing.

This alternative DSP based approach enables the development of more compact and flexible emulators to replace the existing analog and optical counterparts. The programmability feature of these emulators enables the coverage of a large span of fiber line scenarios via a simple modification of algorithm parameters. The proposed scheme uses multiple-input multiple-output (MIMO) FIR filters that are lossless and therefore lend themselves as perfect candidates for emulation of fiber channels suffering from PMD without polarization dependent loss (PDL). The concatenated composition of these filters resembles the continuous-time lumped model of PMD channels and offers a flexible emulator structure in terms of computational complexity which constitutes the main bottleneck for real-time DSP applications.

Without a restriction on the number of filter sections, the construction of discrete-time PMD emulators is straightforward. As it will be demonstrated in Section IV, with a growing number of filter sections, the properties of the PMD emulator converge to those of a real fiber channel (similar to continuous-time models with a large number of birefringent sections). However, because a discrete-time filter with hundreds of sections would be of no practical interest, we investigate methods for approximating the channel properties of a real fiber with a computationally tractable and implementationwise feasible number of filter sections. Therefore, the real challenge in employing FIR MIMO filters for PMD emulation presents itself as a statistical parameter sampling problem. We strive to

The authors are with the Department of Electrical and Electronics Engineering, Koc University, Rumeli Feneri Yolu, 34450 Sariyer-Istanbul, Turkey. (e-mail: [amahmutoglu@ku.edu.tr](mailto:amahmutoglu@ku.edu.tr); [alperdogan@ku.edu.tr](mailto:alperdogan@ku.edu.tr); [aldemir@ku.edu.tr](mailto:aldemir@ku.edu.tr)).

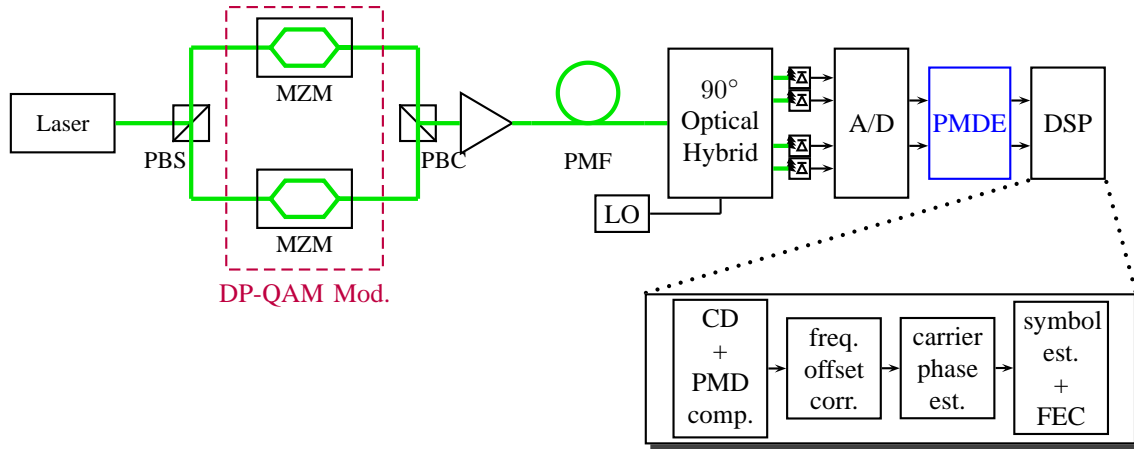


Fig. 1: Outline of an optical communication system with a coherent receiver and the addition of the proposed PMD emulator (PMDE). PBS: polarization beam splitter, MZM: Mach-Zehnder modulator, PBC: polarization beam combiner, LO: local oscillator, A/D: analog to digital converter, DSP: digital signal processor, CD: chromatic dispersion, FEC: forward error recovery [4], [8], [11], [12].

generate a statistical ensemble of filter parameters that produce the correct stochastic behavior of PMD. In the literature, the accuracy of PMD models is usually quantified in terms of the so-called *order* of PMD. The definition of PMD orders is based on the Taylor expansion of the PMD vector around a frequency point [13]. Since the PMD vector has a statistical nature, the individual terms of this expansion have also a random character. However, low order PMD models take only the first few terms of this expansion into account and therefore cannot describe the complex frequency dependent behavior of PMD. For instance, a first-order PMD model uses only the constant term of the Taylor expansion and hence can only describe frequency independent effects. Therefore, these models are only valid for narrowband signals and become increasingly inaccurate as the frequency range of interest grows [14]. In our treatment, we make no assumption on the bandwidth of the signal and take also the frequency dependent behavior of PMD into account. Our methods do not rely on a Taylor expansion around a single frequency. On the contrary, we consider the joint statistical properties of PMD vectors at different frequencies. We strive to produce the correct PMD statistics over the whole frequency range of interest and have the correct frequency autocorrelation [15], [16]. This way we build a PMD emulator that can also be used for the testing of wideband systems.

In the next section, we give a brief outline of the paper and describe the sampling problem at hand without going into details. Sections III and IV are dedicated to the presentation of paraunitary MIMO FIR filters and their properties regarding PMD emulation. In order to address the filter parameter sampling problem, we propose three different techniques and introduce them in Section V in order of decreasing deviation from the desired statistics and increasing computational complexity. Finally, we conclude with simulation results and a brief summary.

## II. FROM ANALOG TO DIGITAL PMD EMULATION

Figure 1 depicts the key components of an optical communication system with a coherent receiver [4], [8], [11], [12]. Note that this back-to-back configuration of the transmitter and the receiver lacks the transmission fiber of a real communication link. A conventional PMD emulator apparatus, usually present between the receiver and the transmitter for testing purposes, is also absent. Instead, the set-up includes a polarization maintaining fiber to connect the transceiver and the receiver. The distortion of the signal for PMD emulation purposes is performed after it is converted to the electrical domain with optical detectors and then sampled with the analog-to-digital converter (A/D). At this stage, the sampled and digitized signal consists of two complex valued components, one for each polarization. In recent high speed experiments with up to 10 Tb/s for a single channel, both of these polarization components at the receiver are modulated using 16-QAM independently [17], [18], which corresponds to a dual polarization quadrature amplitude modulation (DP-QAM) with a 16-ary symbol constellation. Usually, after the A/D conversion stage, the signal is treated with a DSP, compensating for chromatic dispersion (CD) and PMD, and put through a decision circuit after frequency offset correction and carrier phase estimation.

The PMD compensation can be handled by different schemes based on DSP algorithms depending on the requirements of the channel, the modulation method and the transmission speed. For channels, where the pulse broadening due to PMD, i.e. ISI, is negligible, adaptive techniques such as the constant modulus algorithm (CMA) [19] are used for the demultiplexing of the two polarization components [4], [8], [9]. For this task, “a blind source separation scheme based on the magnitude boundedness” of the signals was also proposed in [3]. For higher data rates, an additional compensation of the polarization components needs to be performed in time as well. This can be achieved with methods such as a modification of the CMA algorithm, a fractionally

spaced equalizer (FSE) [1], [20], and algorithms exploiting the paraunitary structure of the PMD channel [21]. In order to evaluate the performance of a PMD compensation method, real life tests are conducted either with actual fibers of realistic lengths [4], [22] or with PMD emulators [7], [8], [9]. Following the example of PMD compensation methods, in order to exploit the advantages of a DSP based emulation technique, we propose to insert a PMD emulation stage between the A/D and the DSP in Figure 1 considering the fact that the signal is already sampled for further processing.

A 2-input 2-output discrete-time filter is required to emulate the effects of a real PMD channel which are of statistical nature. That means we have to choose the filter parameters according to some statistical law that results in accurate channel emulation characteristics such as mean DGD value as well as the whole probability distribution of the DGD and frequency dependent statistics determined by the autocorrelation function (ACF) of the PMD vector [23]. We can use these quantities as performance criteria despite having control only on the parameters of the filter, because we can compute them if we know the transfer function of the filter. Hence, the discrete-time model of a PMD channel can be thought as a generator of PMD vectors with adjustable parameters. For each parameter set, we get a different PMD vector and DGD value at the output.

Similar to the analog case [24], we can construct inherently lossless discrete-time filters for PMD emulation. Each such filter can be decomposed into individual one-tap filter sections (degree-one filters) with DGD values of one symbol period, and conversely, one can construct filters by concatenation of these sections with an additional degree of freedom for the filter length. Although all the sections have the same DGD value of one symbol period, the principal states of polarization (PSP) determined by a single  $2 \times 1$  complex vector parameter for each section is different, and this causes a single state of polarization to propagate at different speeds in each section. The orientations of the PSPs with respect to each other determine the resulting distortion on the signal. Choosing these parameters randomly and in an appropriate manner, we can capture the statistical nature of PMD correctly. This is posed as a problem of random parameter sampling for lossless filters.

Later in Section V we will propose three methods for building discrete time paraunitary models achieving the desired statistical properties. The first parameter sampling scheme, Cascaded Sampling Method (CSM), offers an answer to the following question. How should we sample the filter parameters such that an ensemble of discrete-time filters has an adjustable mean DGD value that is constant over the frequency range of interest? This first step in constructing a discrete-time filter for PMD emulation is important because, as it will be demonstrated in Section IV, when their parameters are not sampled in a special manner, i.e. when they are sampled uniformly and independently, the DGD distribution of lossless filters depends only on the number of sections with a corresponding mean DGD value that grows linearly with the square root of this number. Hence for a fixed filter length and i.i.d. uniform filter parameters, the DGD is not adjustable. Moreover, for filters with a computationally feasible number

of sections, the probability density function (PDF) of the DGD deviates heavily from the desired Maxwellian PDF, especially in the tail region. This deviation is of great importance for the testing of communication systems because the tail of the PDF describes the relative frequency of high DGD values and it is these values that in turn determine the system outage probabilities. Consequently, we want to construct a statistical model parameter sampling scheme that produces Maxwellian DGD PDFs with adjustable means. Another important property we expect is that these statistics are constant over the whole frequency range of interest.

Although we strive for a DGD distribution that is stationary over the frequency range of interest, it proves useful to concentrate on its behavior at the center frequency. It turns out that an instantaneous DGD value at the center frequency has a simple relationship with the filter parameters. This relationship boils down to a sum of random variables problem. Instead of considering the  $N$  filter sections separately and trying to sample their parameters individually, we think of the whole filter as a composition of 2-section blocks. This way we construct building blocks that no longer have a constant DGD. The merging of two degree-one filter sections results in an elementary building block which can also produce fractional DGD values and hence has more expression capacity. This leaves us a greater elbowroom for the design of full length filters in the time domain and sample their parameters accordingly. The parametrization of two-section blocks is, however, different than their one-section counterparts, and since the implementation of paraunitary filters requires the knowledge of the parameters of individual degree-one filter sections, a mapping from the two-section block parametrization to the usual one-section parametrization remains to be found. In the proposed CSM, this mapping is translated into an eigenvalue problem to give the degree-one filter section parameters directly without the need for the costly calculation of transfer functions, resulting in a computationally efficient parameter sampling scheme.

The second method we propose is based on an indirect technique for sampling the filter parameters. This method, Compensated Markov Chain Monte Carlo (C-MCMC), is based on a modified version of the standard Metropolis-Hastings algorithm [25], [26]. This algorithm devises a random walk in the space of filter parameters and the generated samples are the values it assumes during the course of this walk. The random walk can be conceived as a search with an objective or reward function but without a definite target point such as a maximum or a minimum. There are more rewarding, hence more desired points in the filter parameter space since these points produce more probable PMD vector values. At the end, the probability distribution of the ensemble of points visited during the random walk converges to the desired PDF. The random jumps from point A to the point B occur with the help of a decision rule that reflects the weighting of the desired PDF over the space of PMD vectors.

Contrary to the CSM, C-MCMC uses the whole information contained in the PDFs of PMD vectors and not just their Euclidean norms. This enables us to capture the autocorrelation function of PMD vectors over the frequency range of interest

at the expense of increased computational complexity.

In the compensated MCMC scheme, the random walk is constructed in a different space than the decision rule operates in. The decision rule is based on a ratio of PDF values of PMD vectors but we cannot sample these directly. Each set of filter parameters results in a PMD vector at the output but this mapping is neither one-to-one nor isometric. Hence, sampling filter parameters by looking at the corresponding PMD vectors can fail if for example a large number of parameter sets are mapped to the same PMD vector.

The last method we propose is different from the previous two in that it is not probabilistic in nature. Instead we employ a greedy transfer function approximation method (GTFA) [27] in order to capture not only the PMD vector statistics but also the whole transfer function behavior of the channel. We start with the transfer function of a PMD channel generated with a full model. A full model in this context means a PMD channel model with a high number (several hundred to one thousand) of concatenated birefringent sections. Such models are known to accurately approximate PMD channels when their parameters are sampled uniformly and independently. After generating a full model and computing its transfer function, we generate a discrete-time model with a low number of sections and proceed with the greedy approximation. This algorithm optimizes each section of the discrete-time filter locally such that the whole filter with other sections held fixed matches the full model according to some closeness measure. The first part of the local approximation boils down to the closest unitary matrix problem which is easily solved by a singular value decomposition based algorithm. This elegant solution is followed by the second part which uses an eigenvalue equation to finalize the approximation. This procedure is repeated until the desired level of match is obtained. In order to generate another set of filter parameters, this whole routine is repeated. As expected, this method achieves the best results in terms of statistical accuracy of the generated models but it is also the most computationally complex of the three.

All three of these parameter sampling schemes for paraunitary filters are designed to match their PMD vector and DGD statistics to the statistics of a real optical fiber. The goodness of match is measured in terms of the PDFs and their joint moments. A perfect match, in this regard, corresponds to all-order PMD emulation. A partial match, such as in the case of the CSM, represents only the lower orders of PMD. However, one point worthy of notice is that there is no direct connection between the order of moments of the joint PMD distribution on the frequency axis and the order of PMD emulation in the sense it is commonly used in the literature. For example, emulating first-order PMD corresponds to generating instantaneous DGD values that are constant over the frequency range of interest and distributed with a Maxwellian PDF in time. The frequency dependent behavior is captured by higher order PMD. Since all three of the methods we propose produce the required marginal PDFs at every frequency, we are interested in the relationship of our target values at different frequencies. That relationship is given by the statistics of the joint PDFs such as the ACF on the frequency axis. Therefore, matching the behavior of the ACF results in all-order PMD emulation.

On the other hand, a constant ACF amounts to emulating only the first-order PMD. Everything in-between these extreme points emulates higher order PMD to an extent depending on how good the ACF match is. In this sense, all three parameter sampling schemes we propose take all orders of PMD into account and emulate them with a trade-off between accuracy and computational complexity.

### III. DISCRETE TIME LOSSLESS SYSTEMS - THE BASICS

A matrix transfer function  $\mathbf{H}(z)$  is said to be paraunitary if the following holds [28]:

$$\mathbf{H}^*(z^{-*})\mathbf{H}(z) = c^2\mathbf{I}_r \quad \forall z. \quad (1)$$

Here, the superscript “\*” denotes conjugation and transposition,  $\mathbf{I}_r$  is the  $r \times r$  identity matrix and  $c$  is a scalar constant.

If all the entries of  $\mathbf{H}$  are stable transfer functions and equation (1) is satisfied with  $c = 1$ , then  $\mathbf{H}$  is unitary on the unit circle and therefore corresponds to a lossless LTI system. Conversely, it can be shown that for rational transfer functions, the unitariness of  $\mathbf{H}(e^{j\omega})$  for all  $\omega$  implies (1) with  $c = 1$ . Hence a lossless system can be defined as a causal, stable paraunitary system [29]. Since FIR filters are inherently stable, from now on the terms “lossless” and “paraunitary” will be used interchangeably for causal systems.

A general  $M \times M$  degree-one FIR lossless transfer function can be written as [29],

$$\mathbf{H}_i(z) = [\mathbf{I} - \mathbf{v}_i\mathbf{v}_i^* + z^{-1}\mathbf{v}_i\mathbf{v}_i^*]\mathbf{H}_i(1) \quad . \quad (2)$$

Here,  $\mathbf{v}_i$  is a complex  $M \times 1$  vector of unit norm and  $\mathbf{H}_i(1)$  is a unitary matrix. It can easily be verified that this transfer function is paraunitary. The extension to higher degree systems follows naturally with the concatenation of such degree-one blocks and a multiplication with a constant unitary matrix  $\mathbf{R}$ .

$$\mathbf{H}(z) = \mathbf{H}_1(z)\mathbf{H}_2(z)\dots\mathbf{H}_N(z)\mathbf{R} \quad (3)$$

Note that every lossless transfer function can be expressed in this form and this factorization is not unique.

In PMD related calculations we will be dealing with  $2 \times 2$  paraunitary transfer functions that act on  $2 \times 1$  vector valued time series consisting of two orthogonal polarization components of the transmitted signal. In this regard the expression of the corresponding DGD can be obtained based on the expression in (3). The distance between the fast and slow principal states of polarization can be expressed as the difference [30]

$$\tau(\omega) = |\Im\{\lambda_1[\mathbf{G}(e^{j\omega})] - \lambda_2[\mathbf{G}(e^{j\omega})]\}| \quad ,$$

where  $\Im\{\lambda_i[\mathbf{G}(e^{j\omega})]\}$  denotes the imaginary part of the  $i^{\text{th}}$  eigenvalue of  $\mathbf{G}(e^{j\omega})$  and,

$$\mathbf{G}(e^{j\omega}) = \frac{d\mathbf{H}(e^{j\omega})}{d\omega}\mathbf{H}^*(e^{j\omega}). \quad (4)$$

If we now substitute equation (3) in (4) we get

$$\mathbf{G}(e^{j\omega}) = -j \sum_{i=1}^N \prod_{j=1}^i \mathbf{H}_{j-1} \mathbf{v}_i \mathbf{v}_i^* \mathbf{H}_{j-1}^* \quad , \quad (5)$$



with  $\mathbf{H}_0 = \mathbf{I}_2$ . This expression can be greatly simplified for  $\omega = 0$  if we set  $\mathbf{H}_k(1) = \mathbf{I}_2$  for all  $k$ . The DGD of the fiber,  $\tau(\omega)$ , evaluated at the center frequency can in this case be computed to be the difference between the two eigenvalues of the positive definite matrix  $\mathbf{V}\mathbf{V}^*$  with  $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_N]$ . Calculating the eigenvalues of this matrix results in a quadratic equation, and twice the discriminant of this equation gives us the difference between the two eigenvalues.

$$\tau(0) = \sqrt{(a-b)^2 + 4cc^*} \quad (6)$$

where,

$$j\mathbf{G}(1) = \begin{bmatrix} a & c \\ c^* & b \end{bmatrix}, \quad \mathbf{v}_i = \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, \quad \begin{aligned} a &= \sum_{i=1}^N |\alpha_i|^2 \\ b &= \sum_{i=1}^N |\beta_i|^2 \\ c &= \sum_{i=1}^N \alpha_i \beta_i^* \end{aligned} \quad (7)$$

#### IV. DGD STATISTICS OF PARAUNITARY FILTERS WITH UNIFORM I.I.D. PARAMETERS

As a reference point and motivation for further discussion, we investigate the statistical behavior of an ensemble of paraunitary filters when their parameters are sampled uniformly and independently. As mentioned in the introduction, the structure of paraunitary FIR filters as concatenation of degree-one building blocks is similar to the continuous time lumped model of optical fibers [24] and their DGD distribution therefore resembles a real fiber. This resemblance is, however, limited with two important hurdles: The mean DGD of filters tailored with this selection method is fully determined only by the number of degree-one sections,  $N$ , and is not adjustable. Adjustability is a crucial feature for a PMD emulator. Furthermore, the probability density functions deviate heavily from the desired curves particularly in the tails which correspond to high DGD cases and are therefore of primary importance for system outage probability considerations.

The results for the DGD distribution are given in Figure 2 for the uniform i.i.d. parameter sampling method:

$$\alpha_i = \sqrt{X_i} e^{j2\pi Y_i} \quad \beta_i = \sqrt{1 - |\alpha_i|^2} = \sqrt{1 - X_i} \quad ,$$

where  $X_i$  and  $Y_i$  are uniformly distributed in  $[0, 1]$ . The  $X_i$ s constitute a set of  $N$  independent random variables. The same applies for  $Y_i$  and these two sets are also independent from each other. The distribution of the DGD has been computed for different number of sections,  $N = 10$  to  $N = 30$  with  $10^9$  samples and compared with the expected Maxwellian distributions with the same mean as the generated data set. Although the PDF deviation becomes less prominent as  $N$  increases, we cannot rely on this behavior for statistical accuracy since the required number of sections and hence the computational complexity of the resulting emulator would be too high.

Together with equation (6), (7) and the parameter selection method above, the expression for the DGD becomes

$$\begin{aligned} \tau &= \sqrt{(a-b)^2 + 4cc^*} \\ &= \sqrt{(2a-N)^2 + (2\Re\{c\})^2 + (2\Im\{c\})^2} \end{aligned} \quad (8)$$

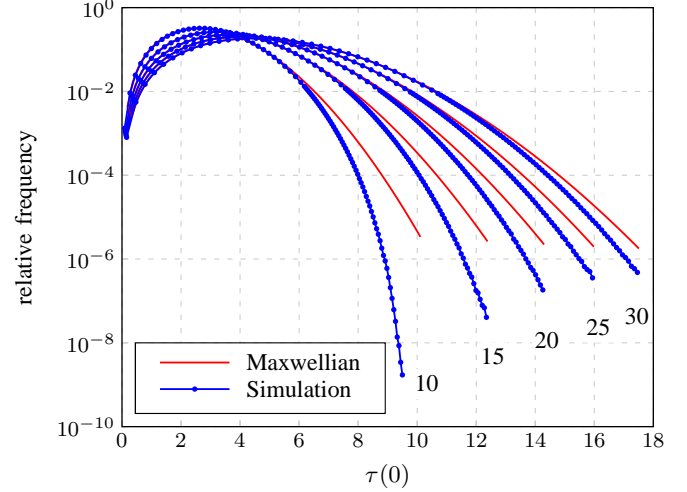


Fig. 2: Probability density functions of the DGD for various number of sections (numbers at the end of the curves) for i.i.d. uniform filter parameters.

$$\begin{aligned} &= \left[ \left( 2 \sum_{i=1}^N X_i - N \right)^2 + \left( 2 \sum_{i=1}^N \sqrt{X_i - X_i^2} \cos(2\pi Y_i) \right)^2 \right. \\ &\quad \left. + \left( 2 \sum_{i=1}^N \sqrt{X_i - X_i^2} \sin(2\pi Y_i) \right)^2 \right]^{\frac{1}{2}} \end{aligned}$$

Remembering that the Maxwellian distribution can be expressed as the square root of sum of squares of three independent zero-mean Gaussian random variables with the same variance, the above expression can be used to verify that the distribution of  $\tau$  indeed approximates a Maxwellian distribution. This is achieved invoking the central limit theorem on each of the three summations of length  $N$ . The first term,  $2 \sum_{i=1}^N X_i - N$ , is zero-mean. Because  $X_i$  are independent, the total random variable  $2a - N$  has the sum of variances of  $X_i$  as its variance.

$$\text{var}(2a - N) = 4N \text{var}(X_i - \frac{1}{2}) = \frac{N}{3}$$

It is again straightforward to calculate the mean and the variance of the second and the third term of the sum. Because  $X_i$  are independent among each other and also independent of  $Y_i$  one can see that  $\Re\{c\}$  is again zero mean because of the symmetry of the PDF of  $\cos(2\pi Y_i)$  around zero. Furthermore the variance of  $2\Re\{c\}$  can be computed as

$$\begin{aligned} \text{var}(2\Re\{c\}) &= N \text{var} \left( 2\sqrt{X_i - X_i^2} \right) \text{var}(\cos(2\pi Y_i)) \\ &= 4N \left( \frac{1}{2} - \frac{1}{3} \right) \frac{1}{2} = \frac{N}{3} \quad . \end{aligned}$$

The same applies for  $2\Im\{c\}$  and hence the total random variable  $\tau$  consists of three approximately normally distributed random variables with  $\mu \approx 0$  and  $\sigma^2 \approx N/3$ . This enables us

to compute the expected mean of  $\tau$  as

$$\bar{\tau} \approx \sqrt{\frac{8}{\pi}}\sigma = \sqrt{\frac{8}{3\pi}}N. \quad (9)$$

These results are valid for the center frequency,  $\omega = 0$  but further investigation reveals that with uniform i.i.d. parameter selection they also hold at other frequencies since there is no bias for any specific frequency point. More in depth discussion of innate statistical properties of PMD emulators with i.i.d. uniform parameters including the derivation of the exact DGD distribution and the examination of the deviation from the Maxwellian is outside of the scope of this paper, but can be found in [31].

## V. PARAMETER SAMPLING METHODS FOR ACCURATE PMD EMULATION

The discussion in Section IV suggests that a different parameter sampling scheme must be implemented in order to achieve accurate DGD statistics. Targeting the DGD statistics is only the first step in the construction of a PMD emulator, since joint PDF statistics for PMD vectors at different frequencies, that reflect the frequency dependent behavior, must be considered as well. To this end, in this section we propose three different parameter sampling methods for discrete time paraunitary FIR filters which can be used for more accurate PMD emulation. From such an emulator we expect a Maxwellian DGD distribution with an adjustable mean value and frequency independent behavior over the whole frequency range of interest. In order to capture higher order effects, we also require a good approximation of the frequency autocorrelation of the PMD vector [23]. With respect to these performance criteria, the three random parameter sampling schemes we propose can be listed and classified as follows:

- *Cascaded Sampling Method (CSM)*: Accurate emulation of first-order PMD with limited control over higher order effects.
- *Compensated Markov Chain Monte Carlo Method (CMCMC)*: Matching the marginal PDFs and the ACF of PMD vectors. Accurate higher order PMD emulation.
- *Greedy Transfer Function Approximation Method (GTFA)*: Transfer function matching and hence accurate all-order PMD emulation.

There is a direct link between this classification and the computational complexity of the sampling schemes: The least “general” method is also the fastest.

Following sections introduce the aforementioned sampling schemes and demonstrate their performance emulating an optical communication link with a bandwidth of 40 GHz and a mean DGD of 0.4 symbol period. All of the experiments use a discrete time filter consisting of 20 birefringent sections. Furthermore we assume that the continuous time system was sampled four times over its minimum rate ( $4 \times 40$  GHz) and therefore set the mean DGD of the emulators to 1.6 sampling periods.

### A. Cascaded Sampling Method

In order to sample the filter parameters in such a way so that the mean DGD becomes adjustable and frequency inde-

pendent, one can extend the degree-one paraunitary building blocks of the filter to 2-section blocks that have dependent  $\mathbf{v}_i$ . We call this technique “cascaded sampling” and analyze its properties below.

1) *Constructing a Paraunitary FIR Filter with Maxwellian DGD Distribution at the Center Frequency*: The difference between the eigenvalues of a Hermitian matrix,

$$\mathbf{G} = \begin{bmatrix} A & C \\ C^* & B \end{bmatrix}, \quad (10)$$

has the form:  $\tau = \sqrt{(A-B)^2 + 4CC^*}$ . Therefore an ensemble of  $2 \times 2$  Hermitian matrices will have a Maxwellian spacing distribution, i.e. the distribution of the difference of its eigenvalues, if the individual components in (10) are distributed as follows:

$$\begin{aligned} A &\sim \mathcal{N}\left(\frac{N}{2}, \sigma\right) & \Re\{C\} &\sim \mathcal{N}(0, \sigma) \\ B &= N - A & \Im\{C\} &\sim \mathcal{N}(0, \sigma) \end{aligned} \quad (11)$$

Here  $\mathcal{N}(\mu, \sigma)$  denotes the Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$ . The resulting Maxwellian distribution will have the following mean value:

$$\bar{\tau} = 2\sqrt{\frac{8}{\pi}}\sigma. \quad (12)$$

At this stage one can ask if unit norm vectors  $\mathbf{v}_i$  can be found such that  $\mathbf{G} = \sum_{i=1}^N \mathbf{v}_i \mathbf{v}_i^*$ . If one can find such vectors they can be used to construct a paraunitary FIR filter so that it will have a Maxwellian DGD at the center frequency by construction. This is indeed the case if  $\text{tr}(\mathbf{G}) \in \mathbb{Z}$  and  $\text{tr}(\mathbf{G}) \geq \text{rank}(\mathbf{G})$  [32].

Since  $N = \text{tr}(\mathbf{G})$  can be restraint to positive integers, this condition can always be satisfied. Being the sum of rank 1 projection matrices,  $\mathbf{v}_i \mathbf{v}_i^*$ , puts one additional constraint on  $\mathbf{G}$ :  $\mathbf{G}$  must be positive definite. This enforces the following inequalities on the parameters in (11).

$$A > 0, \quad B > 0, \quad AB > |C|^2 \quad (13)$$

Because of these constraints, the distributions of  $A$  and  $C$  in (11) must have finite support and hence cannot be actual Gaussians. In Section V-A4 we will show that adjusting the standard deviation of the distributions one can eliminate this discrepancy for all practical purposes.

2) *The case for  $N = 2$* : Since  $\mathbf{G}$  has real eigenvalues and orthogonal eigenvectors if it is constructed as in (10) and (11) with  $N = 2$ , it can be easily verified that  $\mathbf{G}$  can be partitioned in the following way:

$$\mathbf{G} = \mathbf{v}_1 \mathbf{v}_1^* + \mathbf{v}_2 \mathbf{v}_2^*, \quad \mathbf{v}_1^* \mathbf{v}_1 = \mathbf{v}_2^* \mathbf{v}_2 = 1$$

with,

$$\begin{aligned} \mathbf{v}_1 &= \frac{1}{\sqrt{2}} \left( \sqrt{\lambda_1} \mathbf{e}_1 + \sqrt{\lambda_2} \mathbf{e}_2 \right) \\ \mathbf{v}_2 &= \frac{1}{\sqrt{2}} \left( \sqrt{\lambda_1} \mathbf{e}_1 - \sqrt{\lambda_2} \mathbf{e}_2 \right) \end{aligned} \quad (14)$$

where  $\mathbf{e}_i$  are the unit norm orthogonal eigenvectors of  $\mathbf{G}$  with  $\lambda_i$  as the corresponding eigenvalues. In order to get a marginally uniform distribution for magnitude squares of the

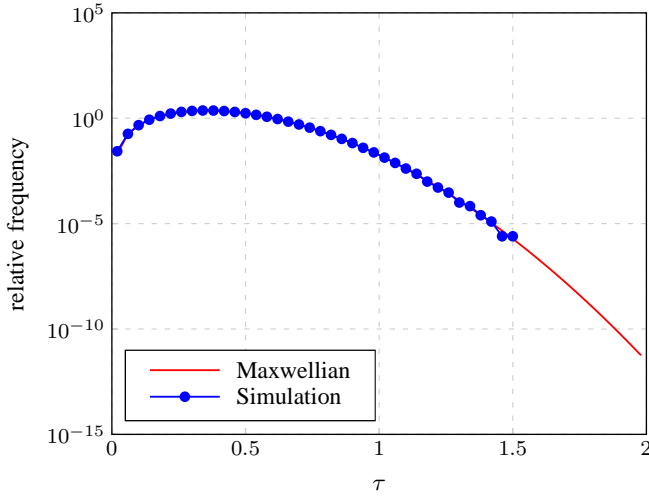


Fig. 3: The PDF of the DGD of the second degree FIR filter compared against a true Maxwellian with the same mean.

components of  $\mathbf{v}_i$ , the eigenvectors  $\mathbf{e}_i$  can be multiplied with a phase term  $e^{j\phi_i}$ , where  $\phi_i$  is selected uniformly at random in  $[0, 2\pi]$ . The resulting empirical DGD density obtained from a simulation with this selection method for  $\mathbf{v}_i$  is displayed in Figure 3. The sample size used in this simulation is  $10^7$  and  $\bar{\tau}$  was taken to be 0.4.

Because of the positive definiteness constraint given in equation (13), the above 2-section unit built with  $\mathbf{v}_1$  and  $\mathbf{v}_2$  can only reach DGD values up to a limit. This shortcoming of the second degree FIR filter can be overcome with the extension of the same idea to higher degree filters. The first possibility that comes to mind is to cascade  $M$  second degree filters and build a more general FIR filter with  $2M$  sections.

3) *Extension to Higher Degree Filters:* Using the fact that the sum of two independent Gaussian random variables is again a Gaussian random variable, we can extend the second degree FIR filter described above to a more general system of even degree.

$$\mathbf{G} = \mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_M$$

$\mathbf{G}$  has a Maxwellian spacing distribution if the summands  $\mathbf{G}_i$  are constructed independently according to (10) and (11). In this case, the unit norm vectors  $\mathbf{v}_i$  can be selected with the same scheme as (14):

$$\mathbf{G} = \sum_{k=1}^M \begin{bmatrix} A_k & C_k \\ C_k^* & B_k \end{bmatrix} = \sum_{k_1, k_2=1}^M \mathbf{v}_{k_1} \mathbf{v}_{k_1}^* + \mathbf{v}_{k_2} \mathbf{v}_{k_2}^* \quad (15)$$

$$\begin{aligned} A_k &\sim \mathcal{N}(1, \sigma) & \Re\{C_k\} &\sim \mathcal{N}(0, \sigma) \\ B_k &= N - a_k & \Im\{C_k\} &\sim \mathcal{N}(0, \sigma) \end{aligned} \quad (16)$$

This filter of degree  $2M$  will have the following mean DGD:

$$\bar{\tau} = 2\sqrt{\frac{8M}{\pi}}\sigma \quad (17)$$

4) *Constraints on Model Parameters:* As discussed earlier, the selection of the standard deviation  $\sigma$  in (11) is not arbitrary.

The fact that  $\mathbf{G}$  must be positive definite restrains  $\sigma$  with an upper bound determined by the minimum probability value we want to match in the DGD distribution. For a fixed value of  $\sigma$  the probability that  $\mathbf{G}$  in (10) is not positive definite can be computed from (13).

The last constraint in (13) takes precedence over the other two since  $A$  and  $B$  can never be negative simultaneously and  $|C|^2$  is always positive. The maximum value  $AB = A(N - A)$  can take is  $\frac{N^2}{4}$ . Hence in the case that  $|C| \geq \frac{N}{2}$ ,  $\mathbf{G}$  is never positive definite. Similarly for  $A \notin (0, N)$   $\mathbf{G}$  cannot be positive definite. This restricts the range of  $\tau$  to  $(0, N)$  as it is expected from an FIR filter built with  $N$  delay elements.

With the conditions above we can calculate the probability that  $\mathbf{G}$  is positive definite,  $P(\mathbf{G} \succ 0)$ , with given  $N$  and  $\sigma$ ,

$$P(\{A = a \mid a \in (0, N)\} \cap \{|C| = c \mid c^2 \leq a(N - a)\}) \quad (18)$$

After a bit of fiddling we arrive at,

$$P(\mathbf{G} \succ 0) = \text{erf}\left(\frac{N}{2\sqrt{2}\sigma}\right) - \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{N^2}{8\sigma^2}}.$$

where  $\text{erf}(\bullet)$  is the error function. Note that this expression is equal to the probability that a Maxwellian random variable with the same mean as in (12) is smaller than  $N$ . In other words, the positive definiteness of  $\mathbf{G}$  ensures that the resulting DGD is at most  $N$  which is the maximum delay of a discrete time filter with  $N$  sections. This result can be used as a filter degree selection tool for a given minimum probability we want to match,  $p$ , and a mean DGD,  $\bar{\tau}$ .

$$\begin{aligned} p &= 1 - P(\mathbf{G} \succ 0) \\ &= \text{erfc}\left(\frac{N}{2\sqrt{2}\sigma}\right) + \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{N^2}{8\sigma^2}} \\ &= 2 \left[ Q\left(\frac{N}{2\sigma}\right) + \frac{N}{\sqrt{2\pi}2\sigma} e^{-\frac{(N/2\sigma)^2}{2}} \right] \end{aligned} \quad (19)$$

where  $Q(\bullet)$  is the Q-function of the standard normal distribution.

5) *Frequency Behavior of the Cascading Method:* The analysis given above only discusses the properties of the cascaded sampling method at the center frequency but the reason this method was chosen among other alternatives to partition a positive definite matrix as sum of idempotents is its uniform frequency behavior.

Figure 4 illustrates this behavior in terms of the PDF of the DGD. These graphics display the DGD distribution of a filter with 20 sections and a mean DGD value of 1.6 first at the center frequency,  $\omega = 0$ , and then at the corner frequency,  $\omega = \frac{\pi}{4}$ . The solid curves in the graphics are the expected Maxwellians with mean 1.6. Figure 4 shows complete agreement of the model output with the desired values at the center as well as the corner frequency. This behavior is typical for the CSM with low enough mean DGD values, however, frequency dependent statistics do not display a similar behavior. Figure 5 shows that although the mean of the DGD distribution is constant over the whole frequency range, the correlation structure deviates from the desired curve.

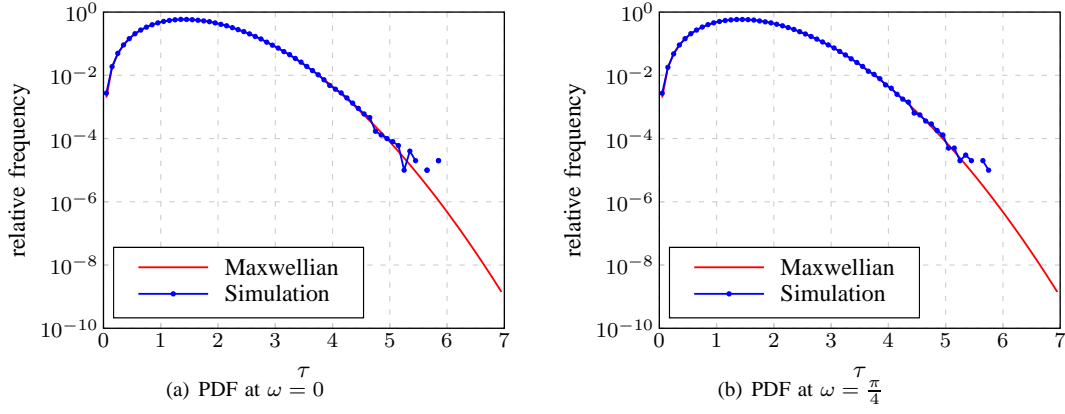


Fig. 4: PDF of the DGD at the center frequency ( $\omega = 0$ ) and the corner frequency ( $\omega = \frac{\pi}{4}$ ) for the CSM. Solid curves represent the Maxwellian with mean 1.6.

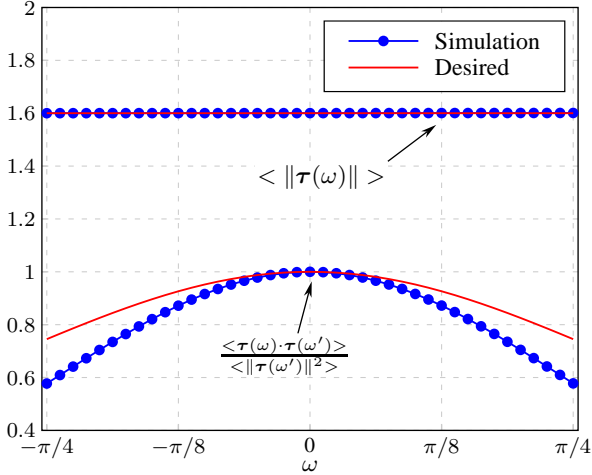


Fig. 5: Mean and normalized autocorrelation curves of the cascaded sampling method compared with the expected values.

Therefore the CSM, in its current form, remains only as a tool for accurately matching the first-order behavior of a real PMD channel.

### B. Compensated MCMC Method

MCMC algorithms are based on the construction of a Markov chain on the sample space of a general vector random variable  $\mathbf{X}$ . In typical settings, the probability density function,  $f_{\mathbf{X}}(\mathbf{x})$ , of the distribution can only be evaluated up to a normalizing constant. The common Metropolis algorithm [25] starts with an initial state,  $\mathbf{x}$ , and generates samples of the random variable iteratively. At every step of the procedure, a new state,  $\mathbf{x}'$ , is proposed according to some proposal distribution  $p(\mathbf{x}, \mathbf{x}')$ . This proposal state is then accepted with a probability determined by the ratio of the PDF values for the new state and the old state.

$$\alpha(\mathbf{x}, \mathbf{x}') = \min \left( 1, \frac{f_{\mathbf{X}}(\mathbf{x}')}{f_{\mathbf{X}}(\mathbf{x})} \right) \quad (20)$$

Because the accept-reject rule only requires the evaluation of the ratio of the probability densities for the proposed and the old state, it is sufficient to know the target PDF up to a scalar constant. The sole restriction of the Metropolis algorithm is that the proposal density be symmetric and simple enough to sample directly.

One generalization of the Metropolis algorithm is the Metropolis-Hastings algorithm [26] which can employ asymmetric proposal densities. In order to achieve this, the acceptance probability is modified so as to incorporate a ratio of the proposal density values.

$$\alpha(\mathbf{x}, \mathbf{x}') = \min \left( 1, \frac{f_{\mathbf{X}}(\mathbf{x}')}{f_{\mathbf{X}}(\mathbf{x})} \frac{p(\mathbf{x}, \mathbf{x}')}{p(\mathbf{x}', \mathbf{x})} \right) \quad (21)$$

Now let us consider how MCMC can be employed to generate samples of fiber models using paraunitary FIR filters. A discrete time fiber model with  $N$  concatenated degree-one sections can be viewed as a complex mapping from the sample space of filter parameters to the space of PMD vector values. This mapping accepts a set of  $2 \times 1$  complex valued unit norm vectors,  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ , which have a total of  $2N$  real scalar parameters as input and produces a frequency dependent PMD vector,  $\boldsymbol{\tau}(\omega)$ , at the output. If we discretize the frequency axis such that we force the statistical properties of the PMD vectors,  $\{\boldsymbol{\tau}(\omega_1), \boldsymbol{\tau}(\omega_2), \dots, \boldsymbol{\tau}(\omega_M)\}$  at a set of frequencies,  $\omega_1$  through  $\omega_M$ , in the frequency range of interest, we can expect the model to behave similarly at intermediate frequencies. Consequently, we obtain a mapping from  $\mathbb{R}^{2N}$  to  $\mathbb{R}^M$ . Therefore, the problem of sampling the input parameters, such that the output statistics exhibit the desired behavior, can be described as follows.

Suppose we are given a general many-to-one, non-isometric map  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$  which maps a random vector  $\mathbf{X}$  to another random vector  $\mathbf{Y} = h(\mathbf{X})$  where  $\mathbf{X} \in \mathbb{R}^n$ ,  $\mathbf{Y} \in \mathbb{R}^m$  and let  $f_{\mathbf{Y}d}$  be the desired probability density of  $\mathbf{Y}$ . Given  $f_{\mathbf{Y}d}$  how must  $f_{\mathbf{X}}$  be chosen such that the transformed variable  $\mathbf{Y} = h(\mathbf{X})$  has the desired PDF?

The answer to this question was given in [33] under the framework of the compensated MCMC algorithm which modifies the accept-reject rule in the standard Metropolis-Hastings



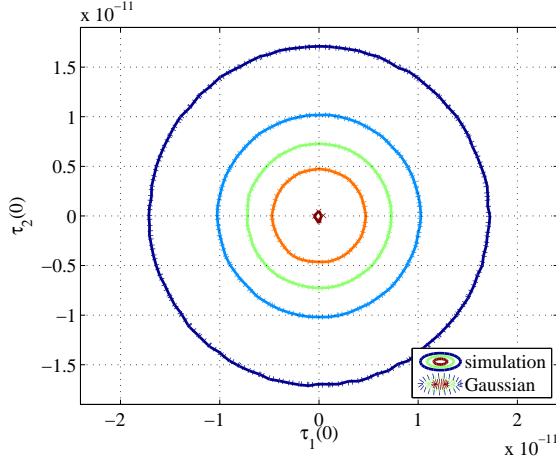
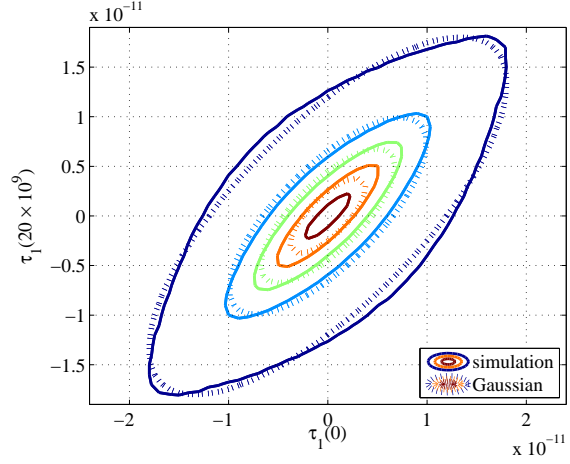
(a) Joint distribution of the first and second components at  $f = 0$ (b) Joint distribution of the first components at  $f = 0$  and  $f = 20$  GHz

Fig. 6: Contour plots of the joint distributions of PMD vector components

algorithm as

$$\alpha_c(\mathbf{x}, \mathbf{x}') = \min \left( 1, \frac{f_{Y_d}(h(\mathbf{x}')) p(\mathbf{x}', \mathbf{x}) f_U(h(\mathbf{x}))}{f_{Y_d}(h(\mathbf{x})) p(\mathbf{x}, \mathbf{x}') f_U(h(\mathbf{x}'))} \right) \quad (22)$$

where  $f_U(\bullet)$  is the distribution of a set of output random vectors consisting of resulting PMD vectors at chosen frequencies when the model parameters are sampled uniformly and independently. This distribution will be called uniform parameter distribution and abbreviated as UPD in the following.

Unlike the cascaded sampling method, where we tried to match only the DGD distribution, with the C-MCMC method we strive to match the frequency dependent statistics of a true PMD channel. In order to achieve this goal, we use the output PMD vector of the model as the target random vector. Due to the complicated nature of the PMD vector, we make use of simplifying assumptions and approximations about its probability distributions concerning the full model as well as the paraunitary FIR filter.

1) *PMD Vector and its Joint PDF at a Particular Frequency*: In the literature, polarization mode dispersion is described in terms of the PMD vector  $\boldsymbol{\tau}(\omega) = \tau(\omega)\mathbf{p}(\omega)$ . Here  $\tau$  is the differential group delay between the fast and slow polarization directions and  $\mathbf{p}$  is the unit vector pointing in the direction of the slow PSP. The PMD vector is defined as the differential change a polarization state at the output of the fiber undergoes as the angular frequency varies.

$$\frac{d\mathbf{s}}{d\omega} = \boldsymbol{\tau} \times \mathbf{s} \quad ,$$

where  $\mathbf{s}$  is a polarization state at the output.

It can be shown theoretically as well as experimentally that  $\boldsymbol{\tau}$  is the result of a random walk process in a three dimensional space and hence has i.i.d. zero-mean Gaussian vector components [30]. As a result of this statistical property, the Euclidean norm of the PMD vector  $\tau$  has a Maxwellian distribution.

2) *Frequency Dependent Behavior of the PMD Vector*: A fiber, subject to polarization mode dispersion, can be modeled

as the concatenation of many statistically independent birefringent sections. As the number of sections grow, the accuracy of the model increases. In order to investigate the joint statistical properties of PMD vectors at different frequencies, a fiber was simulated using 200 sections and  $4.5 \times 10^6$  samples. The mean DGD was set to be 10 psec.

There is no precise theoretical result in the literature about the joint distribution of the components of the PMD vector  $\boldsymbol{\tau}$  at different frequencies. It is known that the marginals of these components are Gaussian and independent at the same frequency. Figure 6(a) shows the contours of the joint distribution of the first and second components of  $\boldsymbol{\tau}$ . The dotted lines are the contours of a jointly Gaussian distribution fitted to the simulation data. In this case, the covariance matrix is

$$C = \begin{bmatrix} 0.3943 & 0.0001 \\ 0.0001 & 0.3921 \end{bmatrix} \times 10^{-22}$$

which is in agreement with the expected result.

On the other hand, one expects the  $i^{\text{th}}$  component of  $\boldsymbol{\tau}$  to be dependent on the  $i^{\text{th}}$  component at another frequency. The joint distribution is again unknown but there is evidence suggesting that it is not jointly Gaussian due to its frequency derivative not being Gaussian [13]. This odd behavior of random variables with Gaussian marginals not being jointly Gaussian can be seen in Figure 6(b). Although there is a clear deviation from the fitted jointly Gaussian distribution, the similarity between two curves hints to an approximation with a jointly Gaussian distribution. The covariance matrix computed for the fitted Gaussian this time is

$$C = \begin{bmatrix} 0.3943 & 0.2951 \\ 0.2951 & 0.3939 \end{bmatrix} \times 10^{-22} \quad .$$

3) *Target Density for Compensated MCMC*: As the discussion in the previous section suggests, we expect an ensemble of PMD vectors to satisfy the frequency dependent statistical requirements of a true PMD channel when the individual components have a jointly Gaussian distribution. Since we assume that only the same components at different frequency

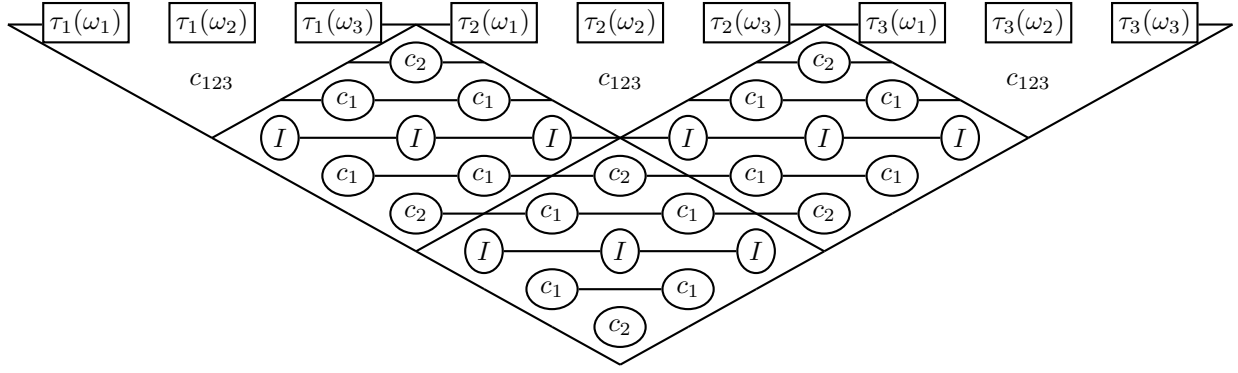


Fig. 7: Copula vine structure for the uniform parameter distribution.

points are dependent on each other, we can express the joint probability density of a collection of PMD vectors as follows:

$$f_T(\mathbf{T}, \bar{\tau}) = K(\Sigma) \exp\left(-\frac{1}{2} \text{tr}(\mathbf{T} \Sigma^{-1} \mathbf{T}^T)\right) \quad (23)$$

Here,  $\mathbf{T}$  is the  $3 \times k$  matrix of PMD vectors,  $\mathbf{T} = [\tau(\omega_1) \tau(\omega_2) \dots \tau(\omega_k)]$ ,  $\Sigma$  is the  $k \times k$  covariance matrix and  $K(\Sigma)$  is the normalizing constant of the PDF.

The covariance matrix depends on only one parameter which is the desired mean DGD,  $\bar{\tau}$ , and can be computed with the expression describing the expected value of the inner product of two PMD vectors at different frequency points  $\omega$  and  $\omega'$  [15]:

$$\langle \tau(\omega) \cdot \tau(\omega') \rangle = \frac{3}{\Delta\omega^2} \left(1 - \exp\left(-\frac{\Delta\omega^2 \langle \bar{\tau}^2 \rangle}{3}\right)\right), \quad (24)$$

where  $\Delta\omega = |\omega - \omega'|$  and  $\langle \bar{\tau}^2 \rangle = \frac{3\pi}{8} \bar{\tau}^2$ . Using (24) we can compute the entries of  $\Sigma$  with

$$\Sigma_{ij} = \frac{1}{3} \langle \tau(\omega_i) \cdot \tau(\omega_j) \rangle, \quad ,$$

since different components of PMD vectors are independent and since the covariance between two components depends on their distance on the frequency axis, this matrix has a Toeplitz structure.

#### 4) Uniform Parameter Distribution of the FIR Filter:

The evaluation of the accept-reject rule in the compensated MCMC algorithm requires the knowledge of the distribution of the PMD vectors when the model parameters are sampled uniformly and independently. Although this distribution resembles the PMD vector distribution of the full model, its exact form deviates from a jointly Gaussian distribution much more than the PDF of the full model does because it has significantly fewer birefringent sections.

Indeed, the effort to approximate this PDF with a joint Gaussian or even a Gaussian mixture model results in inaccurate PMD vector statistics in terms of mean DGD and covariance structure. Moreover, upon close inspection, one can observe that despite being uncorrelated, different PMD vector components at different frequency points exhibit a tail

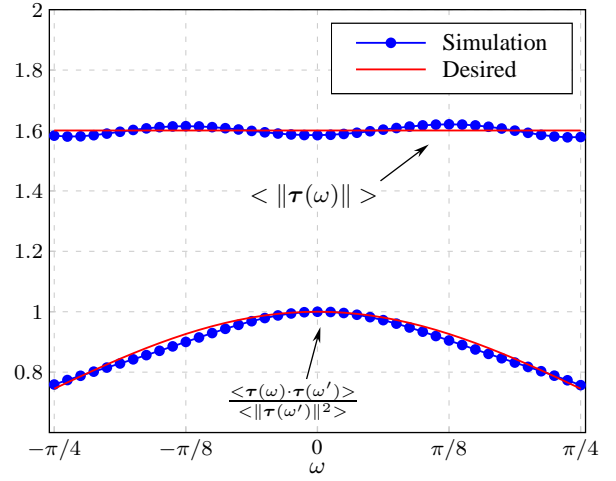


Fig. 8: Mean and normalized autocorrelation curves of the compensated MCMC method compared with the expected values.

dependency which cannot be captured with a jointly Gaussian distribution. In order to overcome these difficulties, we model the uniform parameter distribution as a copula vine [34].

Copulas are multivariate functions that are employed to describe dependency structures of random variables [35], [36]. For our purposes, without going into details, copulas can be viewed as linkage functions that express the joint PDFs of random variables as the product of their marginals and their dependency structure. For the case of two dependent random variables,  $X$  and  $Y$  with marginal CDFs  $F_X$  and  $F_Y$ , one can write

$$f_{XY}(x, y) = f_X(x) f_Y(y) c_{XY}(F_X(x), F_Y(y)) \quad , \quad (25)$$

where  $f_X$  and  $f_Y$  are marginal PDFs of  $X$  and  $Y$  respectively and  $f_{XY}$  is their joint PDF. The non-negative bivariate function  $c_{XY} : [0, 1] \times [0, 1] \rightarrow \mathbb{R}_+$  is the copula density of the two random variables. According to Sklar's theorem [37], under some regularity conditions, such a function always exists. Equation (25) becomes especially useful if the marginals are known. This is indeed the case in the UPD. The PMD vector

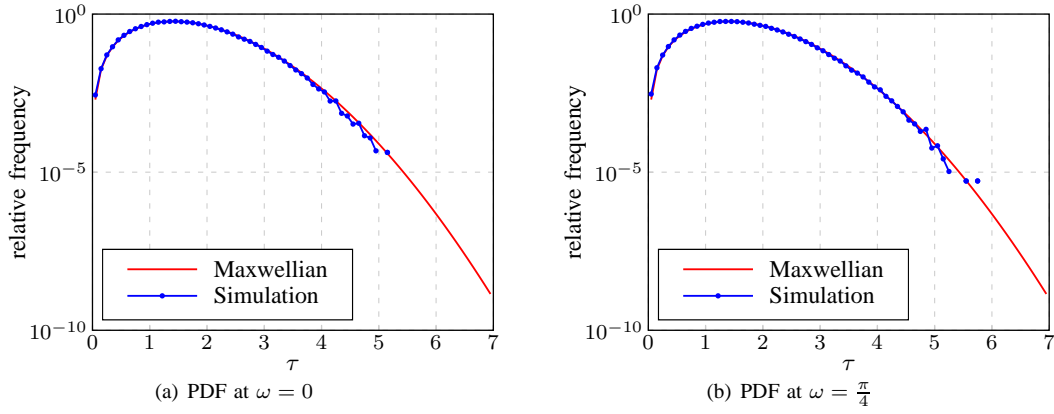


Fig. 9: PDF of the DGD at the center frequency ( $\omega = 0$ ) and the corner frequency ( $\omega = \frac{\pi}{4}$ ) for compensated MCMC method. Solid curves represent the Maxwellian with mean 1.6.

at the  $i^{\text{th}}$  frequency point,  $\tau(\omega_i) = [\tau_1(\omega_i), \tau_2(\omega_i), \tau_3(\omega_i)]^T$ , has independent components that are distributed according to the uniform sum distribution. A random variable  $X$  that is the sum of  $N$  independent uniform random variables has the PDF

$$f_U(x, N) = \frac{1}{2(N-1)!} \sum_{k=0}^N (-1)^k \binom{N}{k} (x-k)^{N-1} \text{sgn}(x-k). \quad (26)$$

Its CDF is given by

$$F_U(x, N) = \frac{1}{N!} \sum_{k=0}^{\lfloor x \rfloor} (-1)^k \binom{N}{k} (x-k)^N. \quad (27)$$

The subscript  $U$  denotes that this distribution is the univariate marginal of the  $\underline{\text{UPD}}$  and  $N$  is the number of sections of the paraunitary FIR filter.

For the construction of the copula density, the multivariate t-copula has proven to be useful [38]. In fact, the set of same PMD vector components (e.g. the first component at all the frequency points) follows a multivariate t-copula distribution with uniform sum marginals almost exactly. The remaining dependency among cross-components is modeled using bivariate t-copulas and arranging them into a D-vine [39] in order to exploit the stationarity property of the joint PMD vector distribution. The parameters of the multivariate as well as the bivariate copulas can be estimated using standard maximum likelihood algorithms.

Figure 7 illustrates the strategy for building trees of pair copulas for three frequency points. The three-variate t-copula,  $c_{123}(F_U(\tau_m(\omega_1)), F_U(\tau_m(\omega_2)), F_U(\tau_m(\omega_3)))$ , is the joint density of the  $m^{\text{th}}$  PMD vector components. The dependency among cross-components,  $(\tau_m(\omega_i), \tau_n(\omega_j))$ ,  $m \neq n$ ,  $i \neq j$ , is accounted for with the pair copulas  $c_1$  and  $c_2$ , and  $I$  is the independence copula that connects the components of a single PMD vector. The symmetry in Figure 7 is due to a simplifying assumption we make in order to obtain a more tractable UPD. Note that the pair copulas connecting the cross-components are in fact conditional PDFs that not only depend on the conditioning variables but also operate on

the transformed forms of their arguments. Here we make the assumption that these copulas are sufficiently flat so that we can describe the relationship among cross-components solely based on their frequency separation. In the end, we obtain the whole joint copula of all nine variables by multiplying all the components in Figure 7.

5) *Performance of Compensated MCMC*: Based on the above, we can construct an accept-reject rule in the compensated MCMC algorithm such that the statistics of the ensemble of output PMD vectors will approximate the desired values. To this end, based on (22), we can write

$$\alpha_c = \min \left( 1, \frac{f_T(\tilde{\mathbf{T}}, \tilde{\tau}) f_U(\mathbf{T}, N)}{f_T(\mathbf{T}, \tilde{\tau}) f_U(\tilde{\mathbf{T}}, N)} \right), \quad (28)$$

The statistics of the C-MCMC algorithm output is illustrated in Figures 8 and 9. The simulation was ran with  $2 \times 10^6$  samples and the first  $10^4$  samples were discarded as the burn-in phase. The number of frequencies in the accept-reject rule,  $k$ , was chosen to be 3 ( $-\pi/4$ , 0 and  $\pi/4$ ). It can be observed that this algorithm performs much better in terms of the autocorrelation function than the cascade algorithm at the cost of a small deviation in the mean value.

The limiting factor on the accuracy of the C-MCMC algorithm is how well the UDP can be modeled and approximated. The simplifying assumptions about the UDP and the pair copulas describing it are sources of inaccuracy. As the number of frequencies in the accept-reject rule, and consequently the number of pair copulas used to approximate the UDP grows, the approximation becomes increasingly inaccurate and the PDFs do not converge to the desired curves. On the other hand, holding the number of frequencies fixed while increasing their distance results in non-uniform frequency behavior of the output. Figure 10 illustrates these two properties together by displaying the mean DGD values for simulations using three and five frequencies over the frequency range  $[-\pi/2, \pi/2]$ , i.e. for a system that is oversampled at twice its minimum sampling rate. The simulation with three frequency points in the accept-reject rule matches the desired mean DGD value at these frequencies but exhibits large deviations at intermediate

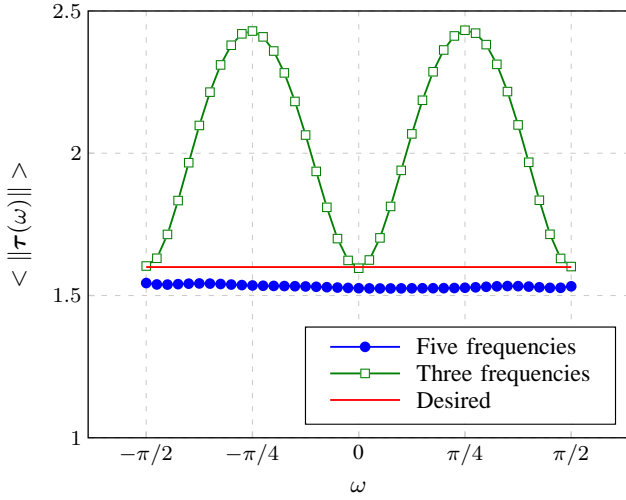


Fig. 10: The mean DGD values for a twice oversampled system with three and five frequency points in the accept-reject rule of the compensated MCMC algorithm.

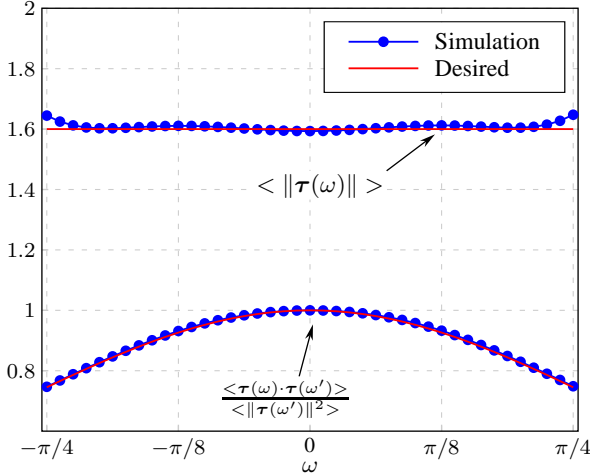


Fig. 11: Mean and normalized autocorrelation curves of the greedy approximation method compared with the expected values.

points while the simulation with five frequencies results in a relatively constant mean DGD value smaller than the desired one.

### C. Greedy Transfer Function Approximation Method

At the heart of the third method we propose for the discrete time PMD emulation parameter sampling problem lies a greedy iterative transfer function approximation algorithm [27]. This algorithm takes a general matrix transfer function as input and tries to approximate it with a paraunitary FIR filter. This is achieved by iteratively optimizing each section of the filter. The GTFA method first defines a distance measure in the space of discrete time transfer functions as a mean-squared weighted Frobenius norm,

$$\xi \triangleq \frac{1}{2\pi} \int_0^{2\pi} W(\omega) \|\mathbf{D}(e^{j\omega}) - \mathbf{H}(e^{j\omega})\|_F^2 d\omega, \quad (29)$$

where, for our purposes,  $\mathbf{D}(e^{j\omega})$  is the transfer function of a real fiber constructed with a high number of birefringent sections (full model) and  $\mathbf{H}(e^{j\omega})$  is the transfer function of the paraunitary FIR filter.  $W(\omega)$  is a weighting function which is set to identity in the frequency range of interest and zero otherwise. This distance is then iteratively minimized by handling the unitary matrix  $\mathbf{R}$  and each degree-one section  $\mathbf{H}_i$  in equation (3) separately.

The optimization of  $\mathbf{R}$  boils down to maximizing  $\Re\{\text{tr}(\mathbf{R}^* \mathbf{A})\}$ , with

$$\mathbf{A} = \frac{1}{2\pi} \int_0^{2\pi} W(\omega) \mathbf{V}^*(e^{j\omega}) \mathbf{D}(e^{j\omega}) d\omega \quad (30)$$

and  $\mathbf{V}(e^{j\omega}) = \prod_{i=N}^1 \mathbf{H}_i(e^{j\omega})$ . This expression can be optimized elegantly by the closest unitary matrix to  $\mathbf{A}$  which in turn can be computed via the SVD of  $\mathbf{A}$  [40]. Similarly the optimization for individual  $\mathbf{v}_i$  in equation (2) is achieved by minimizing the quadratic form  $\mathbf{v}_i^* (\mathbf{B} + \mathbf{B}^*) \mathbf{v}_i = \mathbf{v}_i^* \mathbf{Q} \mathbf{v}_i$  with

$$\mathbf{B} = \frac{1}{2\pi} \int_0^{2\pi} W(\omega) (1 - e^{-j\omega}) \mathcal{R}_i(e^{j\omega}) \mathbf{R} \mathbf{D}^*(e^{j\omega}) \mathcal{L}_i(e^{j\omega}) d\omega.$$

Here,  $\mathcal{R}_i$  and  $\mathcal{L}_i$  denote the right and left factors of  $\mathbf{V}$  respectively, such that  $\mathbf{V}(e^{j\omega}) = \mathcal{L}_i(e^{j\omega}) \mathbf{H}_i(e^{j\omega}) \mathcal{R}_i(e^{j\omega})$ . The minimization of this expression is achieved by setting  $\mathbf{v}_i$  equal to the corresponding normalized eigenvector of the smallest eigenvalue of  $\mathbf{B}$ . It is shown in [27] that with every such iteration the error term in (29) is reduced. Hence, the algorithm moves forward by optimizing each section and the paraunitary matrix iteratively.

The statistical behavior of this method is demonstrated in Figures 11 and 12. It is obvious that the GTFA method outperforms the previous two strategies at the cost of increased computational complexity.

## VI. SUMMARY AND CONCLUSION

We have presented discrete-time PMD models that make use of a paraunitary FIR filter structure to be used in optical coherent receivers. These models can be incorporated into custom chips or off-line data processing devices present in coherent receivers as DSP based PMD emulators for built-in testing. The paraunitary FIR filter structure constitutes a good candidate for PMD emulation not only because of its losslessness property but also for its cascaded nature that enables one to adjust the complexity of the filter. Using three different approaches, we have addressed the question of how the parameters of an ensemble of such filters have to be chosen for them to capture the statistical behavior of a real PMD channel. Using theoretical and simulation results, we have shown that the cascaded sampling method can be employed for



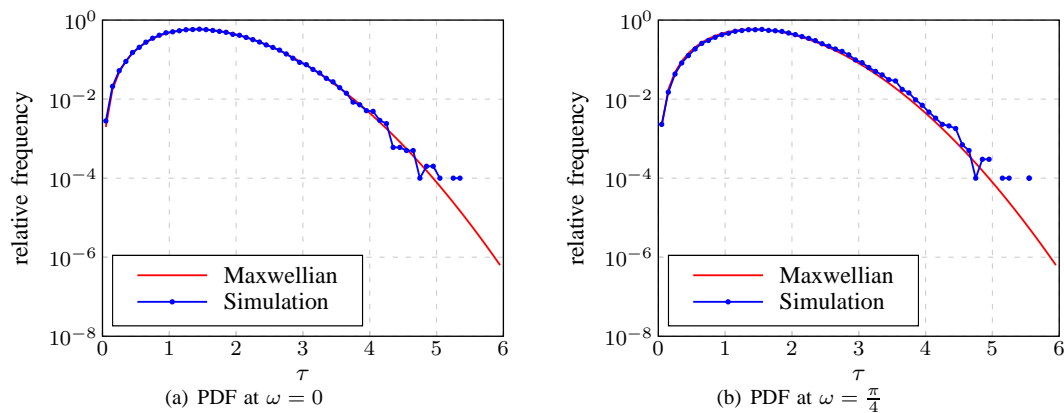


Fig. 12: PDF of the DGD at the center frequency ( $\omega = 0$ ) and the corner frequency ( $\omega = \frac{\pi}{4}$ ) for the greedy approximation method. Solid curves represent the Maxwellian with mean 1.6.

systems which suffer mainly from first-order effects of PMD. For higher order effects, the compensated MCMC algorithm can be used which provides a good approximation for the mean and the autocorrelation values of the PMD vector. The final approach approximates the transfer function of an optical fiber in the frequency range of interest and hence provides the best results in terms of desired statistics. The first two are lower complexity methods which can be implemented in real-time applications while the third one is better suited for off-line signal processing because of its iterative nature. The choice of appropriate method depends on the trade-off between computational cost and the statistical accuracy.

## REFERENCES

- [1] E. Ip and J. Kahn, "Digital equalization of chromatic dispersion and polarization mode dispersion," *Lightwave Technology, Journal of*, vol. 25, no. 8, pp. 2033–2043, 2007.
- [2] A. T. Erdogan, A. Demir, and T. Oktem, "Automatic PMD compensation by unsupervised polarization diversity combining coherent receivers," *Journal of Lightwave Technology*, vol. 26, no. 13, pp. 1823–1834, 2008.
- [3] T. Oktem, A. T. Erdogan, and A. Demir, "Adaptive receiver structures for fiber communication systems employing polarization-division multiplexing," *Lightwave Technology, Journal of*, vol. 27, no. 23, pp. 5394–5404, 2009.
- [4] S. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, pp. 804–817, 2008.
- [5] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, no. 5, pp. 1180–1192, 2010.
- [6] C. Poole and R. Wagner, "Phenomenological approach to polarisation dispersion in long single-mode fibres," *Electronics Letters*, vol. 22, no. 19, pp. 1029–1030, 1986.
- [7] A. Vgenis, C. Petrou, C. Papadakis, I. Roudas, and L. Raptis, "Non-singular constant modulus equalizer for PDM-QPSK coherent optical receivers," *Photonics Technology Letters, IEEE*, vol. 22, no. 1, pp. 45–47, 2010.
- [8] J. Renaudier, G. Charlet, M. Salsi, O. Pardo, H. Mardoyan, P. Tran, and S. Bigo, "Linear fiber impairments mitigation of 40-Gbit/s polarization-multiplexed QPSK by digital processing in a coherent receiver," *Lightwave Technology, Journal of*, vol. 26, no. 1, pp. 36–42, 2008.
- [9] C. Xie, "Polarization-mode-dispersion impairments in 112-Gb/s PDM-QPSK coherent systems," in *Optical Communication (ECOC), 2010 36th European Conference and Exhibition on*. IEEE, 2010, pp. 1–3.
- [10] D. Borne, N. Hecker-Denschlag, G. Khoe, and H. Waardt, "PMD-induced transmission penalties in polarization-multiplexed transmission," *Journal of lightwave technology*, vol. 23, no. 12, p. 4004, 2005.
- [11] L. Nelson, S. Woodward, S. Foo, X. Zhou, M. Feuer, D. Hanson, D. McGhan, H. Sun, M. Moyer, M. Sullivan *et al.*, "Performance of a 46-Gbps dual-polarization QPSK transceiver with real-time coherent equalization over high PMD fiber," *Lightwave Technology, Journal of*, vol. 27, no. 3, pp. 158–167, 2009.
- [12] R. Ludwig, T. Richter, E. Palushani, C. Schmidt-Langhorst, L. Molle, J. Fischer, M. Noelle, R. Elschner, and C. Schubert, "Ultrafast transmission systems using coherent technology," in *Optoelectronics and Communications Conference (OECC), 2011 16th*. IEEE, 2011, pp. 808–811.
- [13] G. Foschini and C. Poole, "Statistical theory of polarization dispersion in single mode fibers," *Lightwave Technology, Journal of*, vol. 9, no. 11, pp. 1439–1456, 1991.
- [14] H. Chen, R. Jopson, and H. Kogelnik, "On the bandwidth of higher-order polarization-mode dispersion: The Taylor series expansion," *Optics Express*, vol. 11, no. 11, pp. 1270–1282, 2003.
- [15] M. Karlsson and J. Brentel, "Autocorrelation function of the polarization-mode dispersion vector," *Optics Letters*, vol. 24, no. 14, pp. 939–941, 1999.
- [16] M. Shtiaf and A. Mecozzi, "Study of the frequency autocorrelation of the differential group delay in fibers with polarization mode dispersion," *Optics Letters*, vol. 25, no. 10, pp. 707–709, 2000.
- [17] T. Richter, E. Palushani, C. Schmidt-Langhorst, M. Nölle, R. Ludwig, and C. Schubert, "Single wavelength channel 10.2 Tb/s TDM-data capacity using 16-QAM and coherent detection," in *Optical Fiber Communication Conference*. Optical Society of America, 2011.
- [18] E. Palushani, C. Schmidt-Langhorst, T. Richter, M. Nölle, R. Ludwig, and C. Schubert, "Transmission of a serial 5.1-Tb/s data signal using 16-QAM and coherent detection," in *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on*. IEEE, 2011, pp. 1–3.
- [19] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *Communications, IEEE Transactions on*, vol. 28, no. 11, pp. 1867–1875, 1980.
- [20] N. Kaneda and A. Leven, "Coherent polarization-division-multiplexed QPSK receiver with fractionally spaced CMA for PMD compensation," *Photonics Technology Letters, IEEE*, vol. 21, no. 4, pp. 203–205, 2009.
- [21] T. Oktem, A. T. Erdogan, and A. Demir, "Adaptive receiver structures for fiber communication systems employing polarization division multiplexing: high symbol rate case," *Journal of Lightwave Technology*, vol. 28, no. 10, pp. 1536–1546, 2010.
- [22] J. Fischer, R. Ludwig, L. Molle, C. Schmidt-Langhorst, C. Leonhardt, A. Matiss, and C. Schubert, "High-speed digital coherent receiver based on parallel optical sampling," *Journal of Lightwave Technology*, vol. 29, no. 4, pp. 378–385, 2011.
- [23] R. Khosravani, I. Lima Jr, P. Ebrahimi, E. Ibragimov, A. Willner, and C. Menyuk, "Time and frequency domain characteristics of polarization-mode dispersion emulators," *Photonics Technology Letters, IEEE*, vol. 13, no. 2, pp. 127–129, 2001.
- [24] A. Demir and A. T. Erdogan, "Emulation and inversion of polarization mode dispersion: A lossless system and reduced-order modeling perspective," *J. Lightw. Technol.*, vol. 26, no. 17, pp. 3071–3089, 2008.
- [25] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller *et al.*,

- "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, p. 1087, 1953.
- [26] W. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
  - [27] A. Tkachenko and P. Vaidyanathan, "Iterative greedy algorithm for solving the fir paraunitary approximation problem," *Signal Processing, IEEE Transactions on*, vol. 54, no. 1, pp. 146–160, 2006.
  - [28] P. P. Vaidyanathan and Z. Doganata, "The role of lossless systems in modern digital signal processing: A tutorial," *Education, IEEE Transactions on*, vol. 32, no. 3, pp. 181–197, 1989.
  - [29] P. P. Vaidyanathan, *Multirate systems and filter banks*. Pearson Education India, 1993.
  - [30] J. Damask, *Polarization optics in telecommunications*. Springer Verlag, 2005, vol. 101.
  - [31] M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *Lightwave Technology, Journal of*, vol. 19, no. 3, pp. 324–331, 2001.
  - [32] R. Hartwig and M. Putcha, "When is a matrix a sum of idempotents?" *Linear and Multilinear Algebra*, vol. 26, no. 4, pp. 279–286, 1990.
  - [33] A. G. Mahmutoglu, A. T. Erdogan, and A. Demir, "Random input sampling for complex models using markov chain monte carlo," *arXiv:1211.4706*, Nov. 2012. [Online]. Available: <http://arxiv.org/abs/1211.4706>
  - [34] T. Bedford and R. Cooke, "Vines—a new graphical model for dependent random variables," *The Annals of Statistics*, vol. 30, no. 4, pp. 1031–1068, 2002.
  - [35] R. Nelsen, *An introduction to copulas*. Springer Verlag, 2006.
  - [36] E. Frees and E. Valdez, "Understanding relationships using copulas," *North American actuarial journal*, vol. 2, no. 1, 1998.
  - [37] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, no. 1, p. 11, 1959.
  - [38] S. Demarta and A. McNeil, "The t copula and related copulas," *International Statistical Review*, vol. 73, no. 1, pp. 111–129, 2005.
  - [39] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 182–198, 2009.
  - [40] R. Horn and C. Johnson, *Matrix analysis*. Cambridge Univ Pr, 1990.